DOCUMENT RESUME

ED 227 177 TM 830 207

AUTHOR Petrosko, Joseph M.; Kapel, David E.
TITLE On Evaluating Controversial Programs: Practical Issues for the School and Community Researcher.
PUB DATE Apr 83
NOTE 26p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, Montreal, Quebec, Canada, April 11-15, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Affective Objectives; *Cognitive Objectives; *Educational Research; Moral Issues; Political Issues; Problem Solving; *Program Evaluation; *Research Problems; Social Problems
IDENTIFIERS *Controversy

ABSTRACT

The nature of controversial programs is discussed, including problems that such programs pose for evaluators and suggested solutions for the problems. Controversial programs are likely to generate news-media attention, internal conflicts (i.e., involving the program and evaluation staffs) and external conflicts (i.e., involving various advocacy groups and program critics). Ways of dealing with these issues are presented. Various experimental and quasi-experimental designs are suggested for evaluating certain types of high-controversy programs. Also discussed are measurement problems facing the evaluator: problems involving the measurement of both cognitive and affective objectives. Among the most salient differences between controversial and non-controversial programs are the factors surrounding the evaluator's decision to evaluate the program. A presentation is made of the negative and positive factors that the evaluator might consider before deciding to evaluate a controversial program. It is concluded that no single set of problem-solving solutions will work for all controversial programs. More work is needed to explicate the evaluation problems inherent in such programs. (Author/PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED227177

On Evaluating Controversial Programs: Practical Issues for the School and Community Researcher

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Joseph M. Petrosko

David E. Kapel

University of Louisville

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. M. Petrosko & D. E. Kapel

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 830 207

Paper presented at the annual meeting of the American Educational Research Association in Montreal, Canada, April 1983.

ABSTRACT

The nature of controversial programs is discussed, including problems such programs pose for evaluators and suggested solutions for the problems. Controversial programs are likely to generate news-media attention, internal conflicts (i.e., involving the program and evaluation staffs) and external conflicts (i.e., involving various advocacy groups and program critics). Ways of dealing with these issues are presented. Various experimental and quasi-experimental designs are suggested for evaluating certain types of high-controversy programs. Also mentioned are measurement problems facing the evaluator: problems involving the measurement of both cognitive and affective objectives. Among the most salient differences between controversial and non-controversial programs are the factors surrounding the evaluator's decision to evaluate the program. A presentation is made of the negative and positive factors that the evaluator might consider before deciding to evaluate a controversial program. It is concluded that no single set of problem-solving solutions will work for all controversial programs. More work is needed to explicate the evaluation problems inherent in such programs.

Among the objectives of this paper are a discussion of the nature of controversial programs and an exploration of how such programs affect the work of the evaluator. Although most of the discussion and most of the examples presented will deal with educational programs, many of the points we raise will be relevant to social service programs in general. Thus, we address concerns that are pertinent to educators, applied social scientists, and most particularly, program evaluators.

The evaluator facing the task of assessing a controversial program would benefit from an analysis of the conflicts and issues involved in such an activity. The literature in social science and evaluation contains discussion of the conflicts and ethical dilemmas facing the applied researcher (American Psychological Association, 1977; Anderson & Ball, 1978; The Joint Committee on Standards for Educational Evaluation, 1981). But little has been written on the problems of evaluating controversial programs per se. Such discussion is needed. The heightened militancy of various advocacy groups has increased the likelihood that educational or social change programs will become associated with controversy. Many of the problems inherent in assessing controversial programs are likely to be encountered sometime in the career of most evaluators.

Characteristics of Controversial Programs

A program can be defined as an organized system of persons, processes, and resources designed to serve human needs. The typical educational program has students (often young people) as recipients of program services, with some combination of cognitive, affective and physical needs being addressed. Some programs deal with students only indirectly, and have as their direct targets teachers,

parents or some other category of persons (Rossi & Freeman, 1982). Parent education programs and teacher inservice would be examples of programs whose ultimate beneficiaries (students) are affected indirectly rather than directly.

Organized system are key words in the definition of program. The formality implied by these words indicates that the program reflects school policy and, at least in some tacit sense, the wishes of policy-makers such as school board members and school administrators. A program is initiated consciously--some person or group has to authorize its implementation. The program has a structure, for example, curriculûm materials, instructional plans, instructors. Some person has superordinate authority over the program, and usually responsibility for program management is vested in another individual or group.

The formal nature of programs is being emphasized because controversial programs will be discussed in the full sense of the term "program". Controversies abound in education, but many are not program related. For example, a particular teacher in a school may use unique instructional techniques that become controversial when publicized. Similarly, personality quirks and idiosyncrasies may make a teacher or administrator controversial, but such cases lie outside the boundary of the present discussion.

Having briefly described the conception of a program, a consideration of the nature controversial programs is necessary. An overall sense of such programs might be gained by examining an arbitrary list of them. Shown below are a set of programs that have generated controversy in recent years. Accompanying each is a brief

description of the nature of the controversy generated by the program.

Program	Area of Controversy/Criticisms
Values clarification (VC)	Not the business of the school, forces students to question their values and often the values found in the home
Sex education/Human sexuality (SE)	Instruction should be directed by parents and/or religious educators; content of courses encourages sexual experimentation and promiscuity
Man-A Course of Study (MACOS)	Teaches that moral values are relative to a particular culture, undermines Judeo-Christian values
Biology (standard curriculum) (BIO)	Ignores creationist theory, implies evolution is a fact rather than a theory
Modern literature (MLIT)	Uses material that contains vulgar, obscene language; presents characters who are immoral, ethnically prejudiced
Modern math (MMATH)	Too theoretical and impractical; fails to emphasize basic computational skills
Career education (CED)	Narrow, does not cover the academic basic skills; dominated by business interests; a form of "sorting" of students
Religion (REL)	Prohibited by separation of church and state
Comparative Government (CGOV)	Invites invidious comparisons between U.S. governmental system and governmental systems of other nations
Citizenship/Patriotism (CIT)	Encourages unthinking, uncritical acceptance of current American society and governmental policies

While the nature of the conflicts differ in these examples, it car be argued that there are common elements among these controversial programs. Most importantly, a value conflict is involved in each of them. Typically, program implementers view the program in question as value neutral, or perhaps "value innocuous." In contrast, critics view it as value-negative, and supporters as value-positive. Critics of the program will justify their objection to it on such grounds as the program is "not the school's business" or that it is "not the basics."

There are probably a number of possible schemes for categorizing controversial programs. One scheme is presented to illuminate two dimensions on which such programs can vary. The model is useful in communicating features of controversial programs that have implications for the evaluator.

First, programs can vary along a dimension of conventionality of curricular content. At one end of this dimension would be programs that would be part of virtually every school curriculum. At the other end would be optional, untypical programs that often would not be found in an average school. Secondly, programs can vary along a dimension of ideology of program critics. For lack of better terminology, the end points of this continuum have been labeled conservative and liberal. In using the latter labels, the authors recognize that these terms are imprecise and that emerging political/social currents such as neo-conservatism (Steinfels, 1979) are rendering the terms less useful than they once were.

Figure 1 illustrates the two dimensional model. Displayed within it are the examples of controversial programs discussed previously.

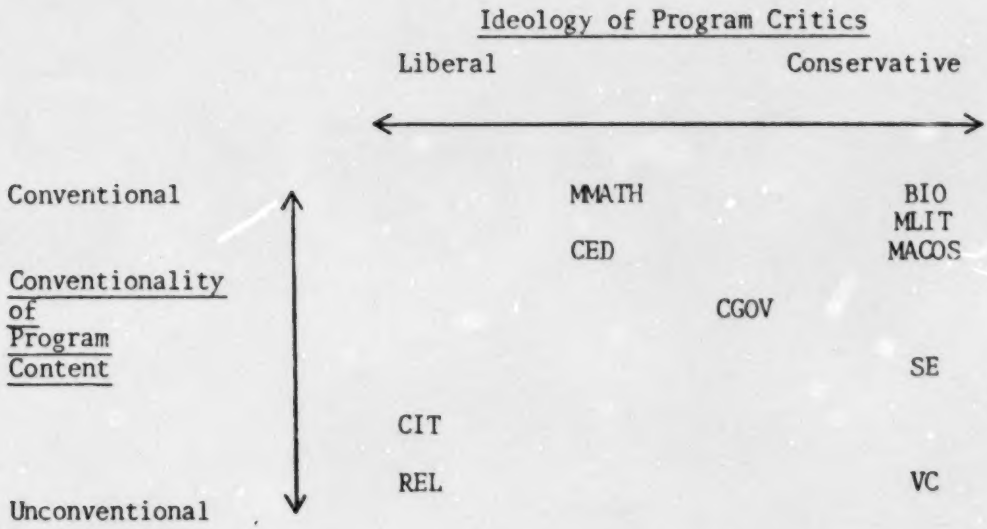

Figure 1. Scheme for categorizing controversial programs

While the precise location of some of the programs in the two dimensional model is debateable, the general location of most of them would be defensible. The model points up two things of which evaluators should be aware. First, controversy can be associated with not just untypical or "offbeat" programs. It can occur with the most standard and accepted subject areas in the curriculum such as mathematics or biology. Secondly, controversy can be generated by critics of all shades of ideology: leftwing, rightwing or middle-of-the-road. Although media attention probably focuses more on conservative critics, liberal or radical criticism of certain programs has occurred. There are also programs where the criticism does not seem to be primarily ideological in nature, or the ideology of critics may vary. For example, modern math was criticized for emphasizing certain content (e.g., set theory) to the detriment of student progress in other content areas (basic computational ability). It is not clear that such criticism was ideologically liberal, conservative, or a mixture of both. It is questionable if the criticism could be said to be ideological at all.

General Issues in Evaluating Controversial Programs

It is probably best to think of programs as varying along a dimension of controversiality rather than as being unambiguously controversial or uncontroversial. If a program, in whatever manner, has become thought of as controversial, what factors are more likely to be an issue for the evaluator of such a program? Some attempts to deal with this question follow. It is assumed throughout this discussion the evaluator is engaged in a summative evaluation of a program--not a formative evaluation of instructional material. The

summative evaluation would typically include outcome measures of any overall program effects.

Internal Conflicts

Controversial programs are more likely to generate internal conflicts than non-controversial programs. Conflicts involving members of the program staff, of the evaluation staff, or of both staffs comprise "internal" conflicts. Conflicts may result from substantive disagreements on evaluation procedures or from personality clashes among individuals. The latter conflicts are more likely due to the extra tension that surrounds the program. The evaluator should be prepared to take actions to defuse explosive situations. Among useful steps would be the scheduling of sessions (perhaps as part of regular staff meetings) involving evaluators and program personnel to talk through conflicts and to develop practical plans to solve problems. The evaluator may want to explore the substantial literature in organizational development research to identify ways of sustaining morale during stressful periods of internal conflict.

External Conflicts

Controversial programs are also likely to have external conflicts: those involving the program staff or evaluation staff and outside groups or individuals. The program is often controversial to begin with because outside groups have raised questions about its legitimacy. Advocacy groups such as the American Civil Liberties Union, the Moral Majority, and local social action groups can become involved in the scrutiny of the program itself and of any evaluation plans and procedures. Not only will

vested interest groups likely be interested in the evaluation, but a larger number of persons from the general public would also be interested in the evaluation because of the controversial nature of the program. In contrast, non-controversial programs are usually of interest, almost exclusively, to those directly involved with the program (or researchers for "academic" reasons).

No foolproof advice can be prescribed either to prevent external conflicts or to resolve them to the satisfaction of all parties. Obviously, attempts at good communication, via public meetings and forums, might clear up rumors and misunderstandings. The evaluator may have to work as a public educator, patiently (and non-technically) explaining the rationale of social science research procedures to laypersons.

News Media Attention

A factor that often accompanies the external conflicts of a controversial program is the special attention of the news media. Journalists, whether of the print or electronic media, are attracted to the newsworthy aspects of a program--such as conflicts at public meetings and verbal disputes between school officials and program opponents. Vested interest groups are becoming more sophisticated in relations with the media and may call news conferences and attend public meetings en masse to present their point of view.

Evaluators must be prepared to speak in public forums and work with journalists to clearly articulate evaluation activities. If the evaluation is being performed by a team of persons, two or more evaluators would be wise to interact with the media, since each person can reinforce and be ready to clarify the comments of others.

If the evaluation is a one-person effort, some thought might be given to preparing, in advance, brief written descriptions of evaluation activities to supplement oral interviews and to ensure that journalists get precise information about the program and its evaluation.

Design and Measurement Issues in Evaluating Controversial Programs

Thus far, the discussion has dealt with general issues pertaining to controversial programs, such as internal conflicts and external conflicts. Would the actual evaluation activities differ between a controversial program and a non-controversial one? There appear to be no unique, specific evaluation procedures that would apply only to controversial programs. Rather, the evaluator would draw upon the array of known social science design and research techniques. However, some comments can be made on plausible situations that would face the evaluator of a controversial program. The logic of the situation would dictate that certain research approaches and measurement options would have greater relevance when the program evaluated is a "hot potato".

Evaluation Planning

As an overall strategy, the evaluator would be wise to develop an evaluation plan with enough flexiblity that if problems develop in getting access to certain data there is still sufficient data for analysis. It pays for the evaluator to have "fall back options" ready if originally planned procedures cannot be carried out. Relevant here is the strategy advocated by Cronbach and associates of launching a series of parallel studies for the same program evaluation rather than investing all measurement resources in one

measure and a single study (Cronbach, Ambron, Dornbusch, Hess, Hornik, Phillips, Walker, & Weiner, 1980).

Instrumentation

The evaluator should avoid overmeasurement of subjects; however, this suggestion requires elaboration. If the advice of Cronbach et al. (1980) is taken and several studies are made of the same project, different instruments might be used with different samples of subjects, thus avoiding having the same persons being subjected to repeated testing. Multiple measures may turn out to be very feasible with some projects. For example, when measuring attitudes and affective objectives, a mix might be attempted of standard (reactive) measures such as questionnaires with non-reactive measures such as running archival records (Webb, Campbell, Schwartz, Sechrest, & Grove, 1981). Since nonreactive measures are not intrusive upon subjects, overmeasurement (in the sense of overuse of persons) is again avoided.

Instruments, whether aimed at cognitive or affective objectives, can become a major source of disputes in controversial programs. Typical problem areas would be measurement of attitudes or opinions in questionnaires or interviews. Affective domain objectives in some programs are equal in importance to cognitive objectives (e.g., sex education) or are primary to the program (e.g., values clarification). The evaluator must be prepared for conflicts regarding the scope of questions, wording of items and confidentiality of data.

While attitudinal or opinion instruments are probably most often sources of problems, they are not alone. Tests in the cognitive

domain can pose problems. Consider sex education programs. The evaluator might face resistance if he or she proposes measuring student knowledge of certain information, for example, questions on the physiology of human reproduction or methods of contraception (Kapel, 1982). If a standard curriculum program like biology becomes controversial, asking questions about sensitive topic areas (evolution) can become an issue. When educational programs have teachers as the program recipients, any type of teacher testing with cognitive instruments can very frequently be a point of conflict, especially with organized teacher groups.

As mentioned earlier, controversial programs are likely to attract the interest of more people than routine educational programs. Accordingly, more people might have involvement in the clearance and approval of instruments. Instruments are often a sore point with program critics because they, by necessity, focus on the objectives of the program. If instruments are designed (as they should be) to register program effects, they become constant tangible reminders of what the program is trying to do. Since critics are not in favor of the program's objectives to begin with, instruments become a convenient target of criticism.

High reliability and validity of instruments are important at all times, but with controversial programs they are especially important. The evaluator would be wise to use proven instruments with already established and reported reliability and validity data. Special purpose questionnaires for ascertaining factual information such as demographic data should be pilot-tested to assure clarity of question content.

Evaluating Unconventional Programs

Controversial programs that cover non-standard curricular topics (note the lower part of Figure 1) are often programs with volunteer participants. If children are program recipients, the parents are really doing the volunteering, in terms of either encouranging or allowing their child's participation. The effect of self-selection is a complicating factor in designing the evaluation and in interpreting evaluation findings. If one simply compares those who get the program (volunteers) with those who don't get the program (non-volunteers) any differences on posttest measures are clouded by differences in the groups due to volunteering (e.g., extra enthusiasm of the volunteers).

Using a version of a true (i.e., randomized) experimental design might be feasible if one can recruit twice as many volunteers as can be initially served by the program. Also adding to the feasiblilty of this approach are two other considerations: a) that the program be relatively short in duration and b) that it be possible to apply the program in several different time periods. Subjects who volunteer for the program but who serve as controls for an initial application of the program can eventually receive the program in another cycle of program delivery.

Shown below is a suggested scheme for measurement of variables and application of the treatment (i.e., program). The design is an adaptation of what Glaser (1973) calls a Prescreened Controlled Experiment. The conventions established by Campbell and Stanley (1966) are followed. The letter R stands for random allocation of subjects, O stands for an occasion of measurement or testing, and X

for the treatment (i.e., the program). A blank between an O and an O or between an R and an O indicates a control condition (e.g., no program or some variation of the program). Generally, each row of letters represents a separate sample of subjects, and the time order is left (earlier) to right (later). If a letter appears in parentheses, the activity indicated by the letter is optional to the design.

non-volunteers	O_1			(O_3)
volunteers	O_2	R	X	O_4
		R		O_5

Here, O_1 and O_2 stand for measures on background or demographic variables (perhaps taken from existing files on subjects). The letters O_3, O_4 and O_5 stand for measures on some dependent variables sensitive to program effects. The design allows one to compare differences between volunteers and non-volunteers on selected variables (O_1 and O_2). The latter data are relevant to external validity (i.e., generalizability). If the O_4 versus O_5 comparison shows evidence in favor of the treatment ($O_4 > O_5$) one can say something like the following: "For those persons who volunteered, the program was effective since a random sample of those who wanted the program and got it were superior to a random sample who wanted the program but did not receive it. The kind of person for whom the program would probably work would be ...[here a discussion of program participant characteristics and O_1 vs. O_2 differences]".

Use of the design requires some favorable circumstances, not the least of which is willingness of the volunteer group to participate

in the experiment and possibly delay receiving the treatment. Probably short-term programs would be most realistic with such a design since the evaluator could assure the control group subjects that they would soon receive the programs after the experiment ended.

One of the persistent problems of evaluation research is the non-applicability of experimental designs for many field studies of program impact. Situations arise making a true experimental design like the one just discussed unfeasible. What if one does not have a short-term program? Or what if one has a program that must be given to all subjects who request it at the same time? A variety of quasi-experimental designs then become possibilities. An extensive literature has arisen on such designs (Campbell & Stanley, 1966; Cook & Campbell, 1979) and it is not the intention of this discussion to cover the topic with any pretense of comprehensiveness. However, there are several designs that have merit for the evaluator of controversial programs.

An adaptation of what Campbell and Stanley (1966) call the separate-sample pretest-posttest design presents useful possibilities for the evaluator. In the adaptation, it is again assumed that volunteering is an important factor to explicitly consider.

Non-volunteers	O_1				(O_3)
Volunteers	O_2	[R	O_4	X	]
		[R		X	O_5]

The assumption here is that everyone will receive the program. A random half of the same treated group receives a pretest (O_4), and a random half a posttest (O_5). If possible, individuals could be matched on one or two key variables, with a random member of each

pair getting the pretest and the other member of the pair getting the posttest. The design is not a perfect one. The internal validity threats of history (events occuring in the environment in addition to X) and maturation (natural changes occuring in subjects at the same time as X) provide rival explanations for any O_4 versus O_5 differences. The history threat can be reduced if the whole design can be repeated a second time. If O_5 exceeds O_4 both times, the argument is less plausible that some factor extraneous to the program caused the O_5 superiority. As with the true experimental design presented earlier, the evaluator can provide data from O_1 and O_2 documenting how volunteers and non-volunteers differ. This information aids in determining the kinds of persons to which any program effects can be generalized.

Although many types of quasi-experimental designs are available to the evaluator of a controversial program, we would suspect that some would be generally unusable due to factors such as their dependence on frequent measurement. For example, the time series design

$O_1 \quad O_2 \quad O_3 \quad O_4 \quad X \quad O_5 \quad O_6 \quad O_7 \quad O_8$

requires frequent testing of the same persons. Unless archival data or some very nonreactive measurement is used, subject responses on measures may show changes because of factors unrelated to the treatment. For example, subjects taking an attitude measure might begin marking the same responses to items at each testing because they infer that the researcher is expecting them to be consistent over time.

The evaluator of a controversial program is at an advantage if the program is relatively short in duration (say, a few months or less) and if the program is planned for repeated administrations. Repeated program cycles mean that, if necessary, weak designs can be repeatedly applied. If program participants consistently outperform non-participants, the evidence mounts for the efficacy of the program. It was stated that the conclusions based on the separate-sample pretest-posttest design gain validity if posttest performance exceeds that of the pretest for each of the two (or more) cycles of program administration. The same logic would hold for a variety of "patched up" designs. Consider an adaption of a design discussed by Tuckman (1978)

Non-volunteers	O_1			(O_8)		(O_9)
Volunteers	O_2	O_3	X	O_4		
		O_5		O_6	X	O_7

Here the program is delivered twice to two different groups. Arguing in favor of the treatment would be $O_4 > O_3$, $O_4 > O_6$, and $O_7 > O_6$. The comparison O_5 versus O_6 provides a useful check on maturation effects. If the O_6 minus O_5 difference does not equal or exceed the O_4 minus O_3 difference, no evidence exists that subjects are naturally improving on the dependent variable measure. Note that two pieces of the design, O_3 X O_4 and O_6 X O_7, are each rather weak single group pretest-posttest designs. But used in tandem, along with O_5 O_6, the overall

design gains validity. Documenting differences between volunteers and non-volunteers is especially important with this design since no random allocation of subjects is performed.

On Evaluating Conventional Programs

Programs that tend to be conventional in content (note upper part of Figure 1) are more likely to enroll a representative cross-section of types of students. Generally speaking, effects of volunteering are probably not as serious as with unconventional programs. Although volunteer effects are not as great, they are likely to exist to some degree and are worth exploring and measuring. The designs discussed in the previous section, where self-selection factors are explicitly measured, could be employed for such purposes.

On Evaluating Programs with Extremist Critics

The more ideologically extreme the program critics (extreme left and right sides of Figure 1) the more likely that the controversy generated around the program is intense. There are several ramifications of this. Program targets are more aware that they in the program, thus raising the probability of Hawthorne effects occuring. Volunteer effects will also act to complicate evaluation comparisons.

Nonreactive measures are especially appropriate when program critics are extreme. If questionnaires or attitude instruments must be employed in the evaluation, ways of minimizing repeated testing of the same persons should be sought. There may be no effective ways of dealing with Hawthorne effects, except to limit generalizations of evaluation findings to those other places and settings where the program is likely to be controversial.

Data Analysis and Reporting

Data analysis for the evaluation of a controversial program would be no different than for a non-controversial program. There are no data analysis procedures unique to controversial programs, just as there are no design or research procedures unique to them. Reporting of the data analysis should be simple and direct. The evaluator must make an effort to avoid obfuscation and to produce clear, effective reporting. Use of charts and graphs help in this effort, but more important is a spirit of honest communication with the reader. The same spirit should animate the reporting of the findings of the evaluation study as a whole, i.e., the conclusions based on the data anlaysis.

There is no implication that the evaluator should strive for honesty when reporting only on a controversial program and should not strive for it with other types of programs. However, when the program is non-controversial, the evaluator may be able to simply present a complex set of findings and deal with any necessary clarifications as needed. Fewer people are interested or involved with a non-controversial program, so questions about an evaluation will probably be fewer. When they do occur, they can be expeditiously handled. But when a program is controversial more people are interested in the evaluation, including members of vested interest groups, critics and supporters. There are a host of persons ready to read an evaluation report and read into it their preconceptions and expectations. The evaluator should aim at straightforward reporting with carefully chosen language. The evaluator must be explicit with no doubt as to the meaning of the

findings. For example, if some findings are uncertain, the evaluator should explicitly state: "we cannot say if there is a relationship between variables A and B." Using such an approach, the evaluator can help minimize the degree of misinterpretation of the findings.

Attempts can be made to anticipate and counteract the tendency of persons to read into a report their own hopes and biases. Most evaluations result in mixed findings. When an evaluator has some positive and some negative findings, but an overall positive or negative effect seems present, all of this must be made explicit and not left to the mind and predjudices of the reader.

The Decision to Evaluate a Controversial Program or 'Why Did I ever Get Mixed Up in This?"

Among the most salient differences between controversial and non-controversial programs are the factors surrounding the evaluators decision to evaluate the program. A relatively innocuous program evaluation has some of the characteristics of a research study: an inquiry aimed at hypothesis-testing (albeit, with some practical, decision oriented consequences). When the program is controversial, the findings of the evaluation are more likely to have sensitive policy and value-related consequences.

Below are possible negative and positive factors that the evaluator might consider before deciding to evaluate a controversial program.

Factors That May Weigh Against Evaluator Involvement (Negative Factors)

One of the first considerations of the evaluator relates to the nature of the program itself. It would not be wise to evaluate a

program that has objectives that reflect values diametrically opposed to one's own personal values. Several writers (e.g., Rossi & Freeman, 1982) explicitly advise the evaluator to avoid such conflicts and thereby avoid a possible compromise of evaluator neutrality.

Obvious questions arise about the motivation of the evaluation. Is there a real interest in objective evaluation? The danger always exists that the evaluator is being used by the evaluation sponsor to produce an evaluation report that will likely come out positive (or negative) regarding the program. There are many ways that an evaluator can be manipulated. For example, are the political forces in favor or against the program being accurately evaluated? Is support for the evaluation adequate, in terms of time, money, accessability to the right people and the right data? An emerging literature (e.g., Anderson & Ball, 1978; The Joint Committee on Standards for Educational Evaluation, 1981) stresses the importance for the evaluator to obtain the necessary freedom and resources to carry out an effective evaluation. The evaluator also needs solid data on the structure of the program. Is the latter accurately presented? Is the history of program development presented as it actually occured? If these questions are not resolved to the satisfaction of the evaluator, the best decision might be one of not becoming involved with the program.

Among a final set of considerations that may argue against evaluator involvement are those related to the negative consequences of risks that must be faced. A controversial program might damage an evaluator's reputation, even if standard research practices are

fo] owed in the study. A certain aura of "guilt by association" may start following the evaluator. Physical risks and mental health risks are also definite factors to be weighed. Evaluating a controversial program is not "business as usual."

Factors That May Weigh In Favor of Evaluator Involvement (Positive Factors)

A controversial program presents challanges (in assessment, measurement, design, data analysis) to the evaluator that are often not available in more humdrum evaluations. It is possible that a partially successful evaluation of a controversial program may add to the stature of the evaluator in the professional community of evaluators and in the general community. Successfully meeting the demands of a difficult job can help the evaluator grow and be recognized as a professional.

The evaluator may have a personal (but objective and open) interest in the topic area addressed by the programs. The evaluation can thus help satisfy the evaluator's curiosity by answering question about the program, e.g., on its effectiveness and it overall impact. It is even possible that evaluators of controversial programs would be paid more than their counterparts in non-controversial programs. Professional and personal risks are much greater in evaluating controversial programs, and greater financial rewards might be necessary to attract competent professionals to do such "hazardous duty."

Final Comments

Some of the issues surrounding controversial programs have been explored in this paper. In addition some of the problems controversial programs pose for evaluators have been investigated.

It is worth pointing out that each controversial program has unique features and no single set of problem-solving solutions will work for all programs. As always, the evaluator must balance responsiveness to various consituencies (evaluation sponsors, members of the general public) with professional integrity. This balancing act is particularly tricky when controversial programs are involved, and the evaluation community needs to become more aware of the complexity facing the person who chooses to function in this area.

Of necessity, the discussion has focused on a specific subset of issues. Many other topics would be worth exploration. For example, qualitative evaluation techniques have gained popularity in recent years. What implications does this technique have for the evaluator of controversial programs? On a related issue, the whole area of the ethical problems of field research is receiving more attention (e.g., de Voss, Zimpher, and Nott, 1982). Evaluators are key professionals who have much to learn and much to contribute to an emerging debate on the limits and potentialities of social science research.

References

American Psychological Association. Ethical standards of psychologists. Washington, D.C., Author, 1981.

Anderson, S. B., & Ball, S. The profession and practice of program evaluation. San Francisco: Jossey-Bass, 1978.

Campbell, D. T., & Stanley, J.C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1966.

Cook, T. D., & Campbell, D. T. Quasi-experimentation: design and analysis issues for field settings. Chicago: Rand McNally, 1979.

Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R. D., Hornik, R. C., Phillips, D.C., Walker, D. F., & Weiner, S. S. Toward reform of program evaluation. San Francisco: Jossey-Bass, 1980.

de Voss, G., Zimpher, N., & Nott, D. Ethics in field work research: a case study. The Urban Review, 1982, 14, 35-46.

Glaser, D. Routinizing evaluation: getting feedback on effectiveness of crime and delinquency programs. (NIMH Monograph, DHEW Publication No. (HSM) 73-9123). Washington, D.C.: U.S. Government Printing Office, 1973.

The Joint Committee on Standards for Educational Evaluation. Standards for evaluations of educational programs, projects, and materials. New York: McGraw-Hill, 1981.

Kapel, D. Assessing a controversial in-service program: an action research approach. Paper presented at the meeting of the American Educational Research Association, New York City, March 1982. (ERIC Document Reproduction Service No. ED 214 971)

Rossi, P. H., & Freeman, H. E. Evaluation: a systematic approach (2nd ed.). Beverly Hills, CA: Sage, 1982.

Steinfels, P. The neoconservatives. New York: Simon and Schuster, 1979.

Tuckman, B. W. Conducting educational research (2nd ed). New York: Harcourt Brace Jovanovich, 1978.

Webb, E. J., Campbell, D. T., Schwartz, R. D., Sechrest, R. D., & Grove, J. B. Nonreactive measures in the social sciences (2nd ed.). Boston: Houghton Mifflin, 1981.